Robust gait identification using Kinect dynamic skeleton data

Noname manuscript No.
(will be inserted by the editor)

Robust gait identification using Kinect dynamic skeleton data

Elena Gianaria · Marco Grangetto

Received: date / Accepted: date

Abstract Gait has been recently proposed as a biometric feature that, with respect to other human characteristics, can be captured at a distance without requiring the collaboration of the observed subject. Therefore, it turns out to be a promising approach for people identification in several scenarios, e.g. access control and forensic applications. In this paper, we propose an automatic gait recognition system based on a set of features acquired using the 3D skeletal tracking provided by the popular Kinect sensor. Gait features are defined in terms of distances between selected sets of joints and their vertical and lateral sway with respect to walking direction. Moreover we do not rely on any geometrical assumptions on the position of the sensor. The effectiveness of the defined gait features is shown in the case of person identification based on supervised classification, using the principal component analysis and the support vector machine. A rich set of experiments is provided in two scenarios: a controlled identification setup and a classical video-surveillance setting, respectively. Moreover, we investigate if gait can be considered invariant over time for an individual, at least in a time interval of few years, by comparing gait samples of several subjects three years apart. Our experimental analysis shows that the proposed method is robust to acquisition settings and achieves very competitive identification accuracy with respect to the state of the art.

Keywords Gait Recognition, Computer Vision, Biometrics, Person Identification, Microsoft Kinect

University of Turin, Computer Science Department,
Corso Svizzera 185, 10149 Turin, Italy
first.lastname@unito.it

1 Introduction

Biometrics deals with the development of statistical and mathematical methods for measuring and analyzing human biological characteristics. It is mainly employed for authentication in security applications, such as in people identification and access control. It permits to exploit the uniqueness of each person for accomplishing the identification task, by recognizing an individual through his/her biometric traits. These traits can be classified in two categories: (i) physical, concerning the intrinsic characteristics of the body, such as fingerprints, iris, retina, hand and face geometry; (ii) behavioral, related to a specific action, for instance voice, handwriting and gait [1].

The features that are considered as the most reliable for person identification, e.g. fingerprint or retina, usually impose severe constraints in terms of acquisition and recognition and require the subject cooperation. Unfortunately, there are many practical un-controlled scenarios, e.g. in a remote surveillance setting, where the subject cooperation cannot be assumed. As an example recent advances in face recognition methods aims at improving the performance in presence of partial occlusions [2], typical of unconstrained surveillance setups. However, a biometric trait like gait is not considered as a valid system for recognizing individuals over time, because the walking style of an individual could change during lifetime, like other behavioral feature [3].

In this paper we focus on *gait* that has the potential to capture both physical characteristics, e.g. height, leg length, and behavioral features. Moreover, we show that gait analysis imposes minimum constraints in terms of subject cooperation. Finally, we investigate if gait can really not be considered a distinctive biometric trait during the years, or rather if is possible to recognize a person comparing his gait sequences with the ones observed few years before.

Gait, defined as "the manner of walking", has been considered as distinctive biometric feature only recently [4,5], but it exhibits several advantages compared to other biometric indexes [6]. In fact, it can be captured at a distance, through a cheap camera, even when the subject is not aware of the monitoring. For these reasons, gait has been proposed for people identification in a surveillance context [7,8] and it is currently applied by forensic science [9–11].

Automatic systems for gait recognition can follow two main approaches [12]: model-free or model-based. Model-free approaches are based on recognition and tracking of the whole human body, or parts of it, from images and define gait features by analyzing the shape of the silhouette and its motion. These methodologies require low computational costs but, being based on 2-D images, are usually not robust to viewpoints and scale variations. On the other side, model-based approaches track body parts in order to construct a 3-D body model. These approaches require more computational power, but can be made view-invariant and scale-independent.

In this paper we propose a model-based gait recognition algorithm that exploits the 3D skeleton model tracking provided by the Micorsoft Kinect sensor. The proposed approach is based on the processing pipeline shown in

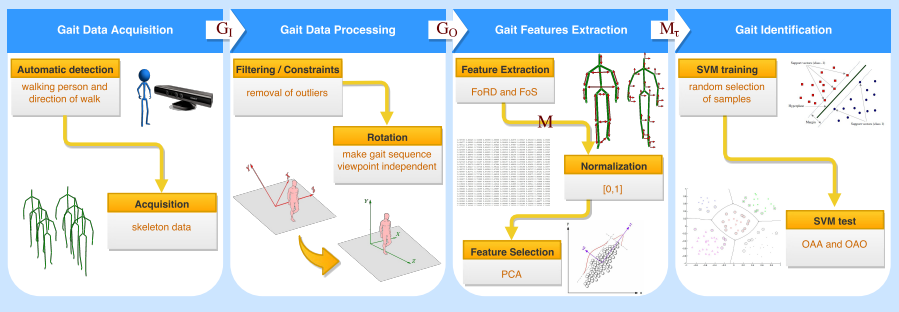


Fig. 1: Overview of our approach

Fig. 1, that includes four phases: *gait data acquisition*, *gait data processing*, *gait features extraction* and *gait identification*. The first step deals with the detection of a walking person and the acquisition of gait data in terms of spatial positions of the skeleton joints tracked by the sensor. In the second step the raw input data are processed in order to remove anomalous and low quality values (when skeleton is not fully tracked or it assumes unnatural positions) and to make them invariant from the sensor viewpoint and the walking path. The third step deals with the extraction of a wide set of features. The last step is devoted to gait recognition and classification using the support vector machine.

The main contributions of the paper are:

- the design of a low-priced identification method based on 3D skeleton tracking, which does not impose specific acquisition settings both on the sensor position and the walking direction;
- the definition and automatic extraction of relevant gait features based on the positions of the skeleton model joints;
- the proposed gait features are determined by taking into account the standard biometric approach that defines the sagittal, transverse and coronal planes as recalled in Fig. 2; these planes are relative to the subject center of mass and depend on the walking direction that is automatically estimated from the tracked joint trajectories. Therefore, the proposed features turns out to be invariant with respect to the sensor position, thus permitting person identification in an un-controlled scenario where the subject is not required to follow any preset trajectory. Clearly, this invariance cannot be guaranteed anymore if one does not respect the acquisition setting imposed by the sensor, e.g. limited field of view and depth range, occlusions, simultaneous tracking of more than two subjects;
- two kinds of gait features are proposed, namely *Features of Relative Distance* between joints (*FoRD*) and *Features of Sway* of joint (*FoS*). The first class represents an estimate of classical physical characteristics, e.g. the length of an arm; the second class conveys information on gait dynamic,

correlated to the behavioral characteristics of the individual. Interestingly, this particular aspect has received limited attention in the literature;
– we accomplish an extensive experimental analysis for validating the proposed approach in terms of identification accuracy. Our experiments have been performed using datasets representative of both controlled and uncontrolled surveillance settings. The obtained results show that the proposed method yield very competitive identification accuracy with respect to recent works in the field.

The rest of this paper is organized as follows: Section 2 presents an overview of the related work on gait recognition systems; in Section 3 the proposed gait acquisition method is described whereas in Section 4 the proposed gait features are defined and applied to the identification problem using principal component analysis and support vector machine. Section 5 presents the datasets of gait sequences we have employed in our study. Experimental analysis is worked out in Section 6 and the conclusions of our work are drawn in Section 7.

2 Related Work

Gait recognition is a recent and open research problem [13]. As mentioned above, gait analysis is mainly carried out by following model-free or model-based strategies. Before the advent of RGB video and depth (RGBD) sensors, the most common approaches were the model-free ones. Two representative studies are those of Wang *et al.* [14], where background subtraction and image segmentation are combined to isolate the spatial silhouettes of a walking person, and of Han *et al.* [15], that have proposed a spatio-temporal gait representation, termed Gait Energy Image (GEI), that represents human motion sequence in a single image while preserving temporal information. More recently, Kusakunniran [16] has presented a new method to extract and recognize spatio-temporal gait features from video sequences directly, without pre-processing. Muramatsu *et al.* [17] and Connie *et al.* [18] have proposed new methods for solving the known view variation problem in the cross-view gait recognition systems, with a view transformation model approach and a Grassmannian approach, respectively.

Also model-based approaches have been proposed. Urtasun *et al.* [19] have defined and extracted the style parameters of the 3-D motion of a whole gait sequence. Bouchrika *et al.* [20] have proposed to extract human joints and construct motion templates for describing gait, based on elliptic Fourier descriptors. Tafazzoli *et al.* [21] have studied the movement of legs and arms in order to construct a body model according to the anatomical proportions. They have described the motion patterns by means of Fourier analysis. Jung *et al.* [22] have analyzed the 3-D gait trajectory to enable face acquisition and recognition in surveillance environments. Zhang *et al.* [23] have proposed a gait recognition system based on wearable accelerometers and portable smart devices, which overcomes the typical limitations of this kind of sensors.

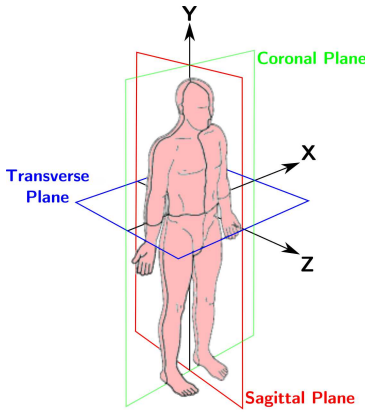


Fig. 2: Human body planes.

The widespread diffusion of gaming peripherals for body tracking based on RGBD sensors, like popular Microsoft Kinect, have given a significant impulse to the study of gait with model-based approaches. However, many of the proposed methods require specific acquisition settings, like walking along a predetermined path. Preis *et al.* [24] have tested different combination of anthropometric features, such as height, length of limbs, stride length and walking speed, for gait characterization. Ahmed *et al.* [25] have employed the new Kinect sensor for extracting two kinds of features, namely joint relative distance (JRD) and joint relative angle (JRA), proved to be robust against view and pose variations. Chattopadhyay *et al.* [26,27] have proposed a gait recognition approach in presence of occlusions using 2 Kinect cameras placed in the entry/exit points of surveillance zones. The depth map is used for aligning the silhouette and make its orientation parallel to the XY plane, while the gait features, related to the movement of lower limb region, are extracted from the skeleton data. Andersson *et al.* [28] and Yang *et al.* [29] have conducted their experiment on the same dataset. In their configuration the subjects have walked on a fixed path and the Kinect sensor is placed on a spinning dish for keeping the subject always in the center of its view. In particular, Andersson *et al.* [28] have calculated both gait and anthropometric attributes. The former are related to kinematic parameters such as angles between joints of lower body part, and spatio-temporal parameters, such as gait cycle size and speed. Yang *et al.* [29] have based their classification on relative distance gait features, and they have improved the classification accuracy using random subspace method for feature selection.

Among the few works that do not rely on any particular acquisition setting we mention Satta *et al.* [30] and Pala *et al.* [31]: in this case anthropometric measures have been combined with clothing appearance to improve re-identification task in different cameras. Kinect skeletal information is used to separate the body silhouette into 4 parts and to extract the relative distances

between joints. Clothing appearance descriptors, based on color of dresses and body, are extracted from the silhouette. Clearly, the color based recognition is applicable only for re-identification system during the same day (for example re-identification in public area in companies, universities, and so on).

Some of the concepts proposed in this paper have been initially explored in our recent works [32–34]. In particular, in [32] we targeted a simpler scenario where gait is used to discriminate only between a pair of subjects with similar biometric features. In [33] some gait features based on Kinect skeletal joint model are preliminarily defined and used for identification in the particular case of a controlled setting where the subject follows a straight path in front of the sensor. In [34] we preliminarily investigate if gait can be considered invariant over time by comparing gait samples of several subjects acquired over a period of three years.

3 Gait Data Acquisition and Processing

Our study focuses on the analysis of gait as distinctive biometric trait used for individual recognition. Since we target a non intrusive and inexpensive tracking system, we select the popular Microsoft Kinect device as motion acquisition sensor. Kinect is an RBGD camera that has revolutionized the Human Computer Interaction (HCI) field. Originally conceived for the gaming industry, its impact has extended far beyond it. Researchers and practitioners are leveraging the technology for application in computer science, robotics, electronic engineering, and medicine [35]. The device is equipped by an RGB camera, a depth sensor and a multi-array microphone. The viewing angle spans 43^o vertically and 57^o horizontally. Moreover, a tilt motor allows an additional $\pm 27^o$ on the vertical field of view. Kinect can capture color, depth data and audio, and process the depth data to generate skeleton data.

Here we exploit the sensor *Skeletal Tracking* capability that allows to track a multi-joints body model in real-time. The Kinect system has several advantages: it does not require any camera calibration or environment setup; it allows marker-less tracking; it is widespread diffused and quite cheap; it provides libraries for creating custom applications.

3.1 Notation

The Kinect skeleton map consists in 20 joints, labeled $J_0, \ldots, J_{19}$, as shown in Fig. 3; each joint is represented by a 3D point in the coordinates system (x, y, z) centered on the sensor, where x and y represent the horizontal and vertical direction respectively, while z represents the depth direction with respect to the optical center. The sensor also provides an estimate of the floor plane where the user is walking. This plane is returned in terms of its normal vector $\mathbf{n} = (a, b, c)$ and the height d of the camera center with respect to the floor: the plane equation can be written in implicit form as $ax + by + cz + d = 0$.

In the rest of the paper we will use the following notation:

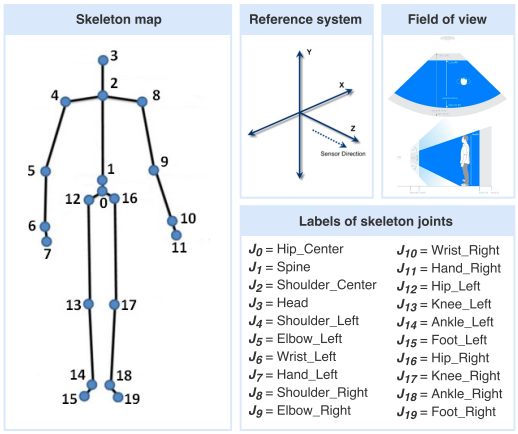


Fig. 3: Details on the Kinect sensor and the skeleton data.

- N: the total number of skeletons estimated (in time) from a video sequence;
- $J^i_k = (J^i_{k,x}, J^i_{k,y}, J^i_{k,z})$: the coordinates of the k-th joints at time i, with $0 \leq k \leq 19$ and $i < N$;
- $d(J^i_h, J^i_k)$: the Euclidean distance between the h-th and the k-th joint;
- Ξ_i: $\left\{J^i_k : 0 \leq k \leq 19\right\}$ the skeleton map at time i.

3.2 Skeleton Acquisition

We have developed an application for collecting gait samples. The software tracks a subject in the scene and automatically recognize if he/she is walking, using the stride detection function described in Algorithm 1. The mechanism is based on the detection of left (and right) foot movements. The foot is considered in movement if its position has changed (with a certain tolerance d_{tol}) with respect to its position in previous Δ_w frames. In this case a binary flag $isWalking$ is set at `true`. If both feet are stable, or the movement is less than d_{tol}, we can assume that there is no movement and $isWalking$ is set to `false`. Our experimentation has shown that Kinect feet tracking can be very defective because of self occlusions and floor/foot segmentation errors. As a consequence, the movement of feet has been inferred in Algorithm 1 from the ankles joints (J_{14}, J_{18}) that have shown to provide more robust results. Based on our experimental analysis we set $\Delta_w = 3$ and $d_{tol} = 5$cm.

When a walking subject is detected the set of 3D joint coordinates (i.e. the skeleton map Ξ_i) is acquired, frame by frame.

For gait identification it is important to separate the features associated to the right and left parts of the body. As a consequence, when the observed subject is wandering freely in the surveilled area it is crucial to discriminate between the cases when he/she is walking toward or away from the sensor. To this end we have implemented a simple mechanism based on the position of

Algorithm 1 Is_Walking

for each frame i **do**
  if $d(J^i_{14}, J^{i-\Delta_w}_{14}) > d_{tol}$ **or** $d(J^i_{18}, J^{i-\Delta_w}_{18}) > d_{tol}$ **then**
    if not $isWalking$ **then**
      $isWalking \leftarrow$ `TRUE`;
    end if
  else
    if $isWalking$ **then**
      $isWalking \leftarrow$ `FALSE`;
    end if
  end if
end for
return $isWalking$

the center of mass (J_0) and we define a binary flag $isFront$ that we set to `true` when the subject is moving toward the sensor and `false` when he/she moving away from the camera. Let us consider the time series $J^i_{0,z}$: if the depth coordinate (z) is strictly decreasing it means that the person is walking towards the camera, i.e. the direction is "front" and we set $isFront = true$. Otherwise, if the value of $J^i_{0,z}$ strictly increases, the person walks far away from the camera, i.e. the direction is "rear" and we set $isFront = false$. In practice, we compare the joint position at time i and at time $i - \Delta_f$ ($J^i_{0,z} < J^{i-\Delta_f}_{0,z}$). In our experiments we used $\Delta_f = 2$. The details of this procedure are given in Algorithm 2. During the real-time execution, when a change of direction is detected, the involved joints (i.e. those of arms and legs) are properly labeled, in order to keep the interpretation of "left" and "right" consistent with the frontal view assumed in Fig. 3.

Algorithm 2 Is_Front

for each frame i **do**
  if $J^i_{0,z} < J^{i-\Delta_f}_{0,z}$ **then**
    $isFront \leftarrow$ `TRUE`;
  else
    $isFront \leftarrow$ `FALSE`;
  end if
end for
return $isFront$

Of course, if the person walks backwards but facing the camera, our algorithm fails. Within the scope of this paper we assumed that this is not applicable. A possible solution to overcome this problem may be to couple our proposed mechanism with a face detection procedure, by processing the RGB stream data as well.

The output of this first phase is a gait sequence G_I, composed by all the collected Ξ_i, which represent the input of the next processing phase.

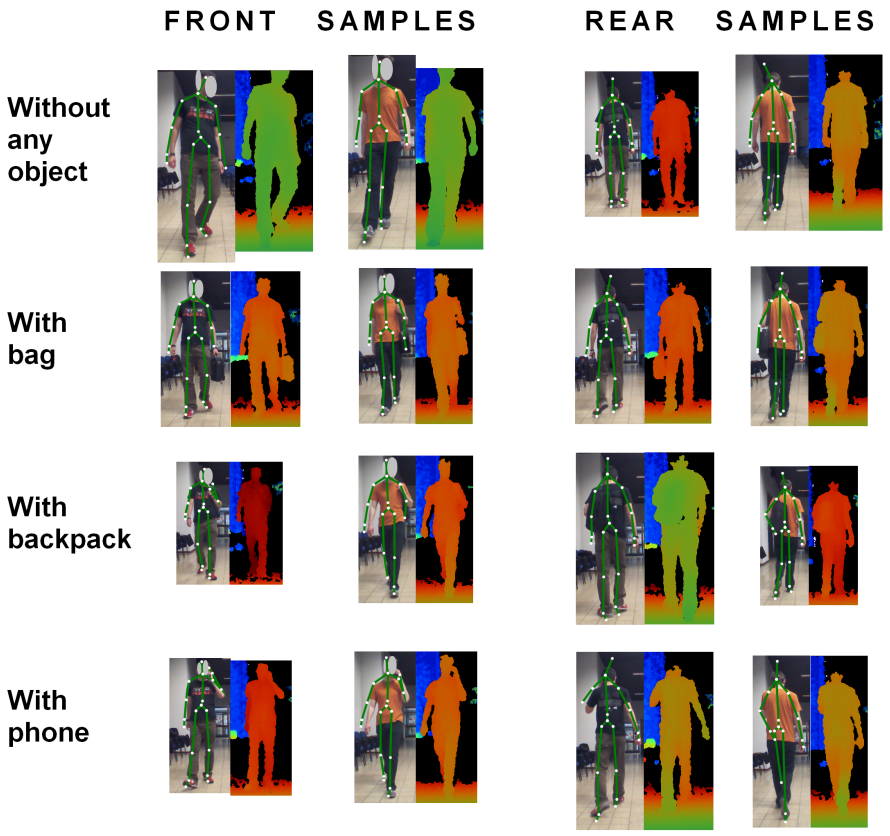


Fig. 4: Examples of tracked skeletons (superimposed on RGB image along with the corresponding depth map) acquired in our surveillance setting.

3.3 Gait Data Processing

The Kinect skeletal tracking is very effective in a standard gaming setup, where the user stand in front of the sensor, whereas in our un-controlled scenario many occlusions due to rear and side poses can lower the precision of the estimate. In Fig. 4 we show some examples of tracked skeletons (superimposed on RGB image along with the corresponding depth map) acquired in our surveillance setting. The images refer to different acquisition cases that will be analyzed in Section 6, namely frontal/rear views and effects of partial body occlusions when subject is carrying an accessory, e.g. a shoulder bag, a backpack or using a phone. It can be noted that skeleton is reliably tracked in frontal views in absence of occlusions (top-left images), whereas in the other cases some joints can be misplaced. Moreover, the sensor is sensitive to the lighting condition and to the color of clothes [36,37]. As a consequence, in our surveillance case the skeletal data are potentially affected by measurement er-

rors. To counteract sensor limitations as much as possible we have designed a set of consistency checks that allowed us to filter out measurements that are very likely to be unreliable.

First of all, we recall that Kinect provides, for each joint of the skeleton model, a tracking state: it can be `tracked` for a clearly visible joint, `inferred` when a joint is not clearly visible and the sensor infers its position, or `non-tracked`, if the joint is outside the field of view. Clearly, only the joints in the `tracked` state have accurate position in space.

A preliminary consistency check is thought for limiting the acquisition errors and select only the really informative frames, i.e. those frames in which the whole skeleton is tracked and each joints is visible. This condition is implemented by keeping in time series G_I only those skeleton map Ξ_i where both head (J_3) and feet joints (J_{15}, J_{19}) are in tracked state. The rationale behind this heuristic is that we can assume that if head and feet are visible at the same time, the entire skeleton is visible.

Another set of consistency checks can be enforced considering the position of left (and right) body joints with respect to the center of mass. Except in limit cases, where the left and right parts of the skeleton occludes each other, if the subject walk toward the sensor, the values for the horizontal position (axis x) of the left joints should be lower than $J^i_{0,x}$, and the opposite for the right parts of the body. The same assumption holds also for rear poses when using the re-labeling of left and right joints according to the output of Algorithm 2. As a consequence, we can enforce the following constraint on the joints bounded to skeleton, i.e. shoulders or hips:

$$\begin{aligned}\chi^i_x =&(J^i_{4,x} < J^i_{0,x}) \wedge (J^i_{8,x} > J^i_{0,x}) \\ &\wedge (J^i_{12,x} < J^i_{0,x}) \wedge (J^i_{16,x} > J^i_{0,x}).\end{aligned}$$

In the same manner, if we consider the vertical direction (axis y) we know that the y coordinate of the lower body parts (knees, ankles and feet) should be lower than $J^i_{0,y}$. Therefore, we can define an additional constraint χ^i_y as:

$$\begin{aligned}\chi^i_y = &(J^i_{13,y} < J^i_{0,y}) \wedge (J^i_{14,y} < J^i_{0,y}) \wedge (J^i_{15,y} < J^i_{0,y}) \\ &\wedge (J^i_{17,y} < J^i_{0,y}) \wedge (J^i_{18,y} < J^i_{0,y}) \wedge (J^i_{19,y} < J^i_{0,y}).\end{aligned}$$

All frames in G_I where such constraints are not satisfied can be dropped from further analysis.

The next processing step represents a key element allowing us to achieve gait identification in an uncontrolled setting and aims at making the acquired gait samples invariant with respect to the mutual position between the sensor and the target subject. In other words, the objective is to obtain gait samples independent from the sensor view point, on the one hand, and from the walking trajectory followed by the subject, on the other hand. To this end we propose a proper transformation into a new coordinate reference system oriented according to the walking direction.

Let us consider again the three basic anatomical planes shown in Fig. 2, namely the sagittal, transverse and coronal planes. Our goal is to transform

the Kinect coordinate system (x, y, z) according to the reference axes (X, Y, Z) that represent horizontal, vertical and walking directions with origin in the center of mass of the subject. As shown in Fig. 2, the planes YZ, XZ and XY represent the sagittal, transverse and coronal planes, respectively. In the new coordinate system, axis Z represents the walking direction, that may change from point to point. Moreover, the transverse plane XZ is parallel to the floor; it follows that the Y axis coincides with the normal vector $\mathbf{n}$ estimated by Kinect.

The only information we need to estimate for computing the coordinate transformation is the instantaneous walking direction. Fig. 5 depicts a sample of the trajectory of the center of mass J_0 projected on the zx plane (Kinect coordinates): the walking direction can be defined as the tangent line to the trajectory of J_0 (time series of the pairs $(J^i_{0,x}, J^i_{0,z})$). By approximating the derivative with the incremental ratio we can estimate the i-th walking direction angle θ_i as:

$$\theta_i = \tan^{-1}\left(\frac{J^{i+1}_{0,x} - J^i_{0,x}}{J^{i+1}_{0,z} - J^i_{0,z}}\right). \tag{1}$$

Now we can define the local system of coordinates by first translating the origin to J_0, and then rotating the axes according to the walking direction and the floor normal. This can be achieved by introducing homogeneous coordinates and defining the following 3×4 matrix transformation:

$$\begin{bmatrix} J^i_{k,X} \\ J^i_{k,Y} \\ J^i_{k,Z} \end{bmatrix} = \mathbf{T}_i \begin{bmatrix} J^i_{k,x} \\ J^i_{k,y} \\ J^i_{k,z} \\ 1 \end{bmatrix} \tag{2}$$

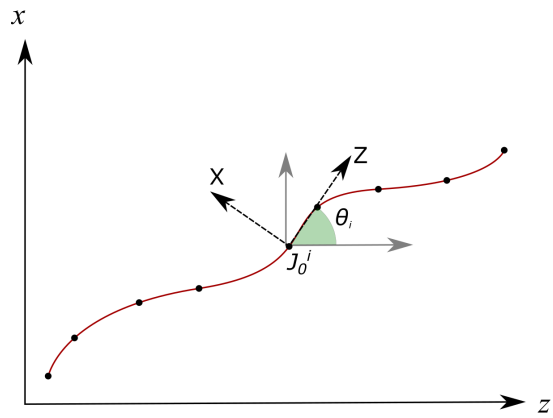
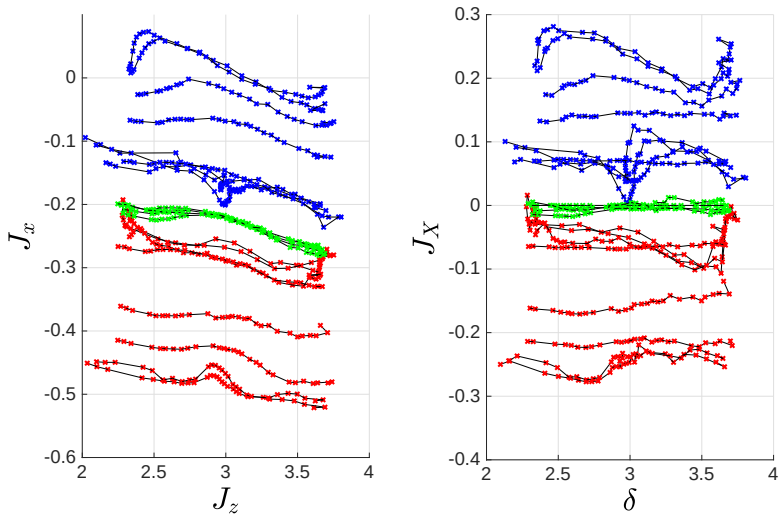


Fig. 5: A sample of trajectory of the center of mass on the zx plane.



(a) (b)

Fig. 6: Top view projection (zx plane) of original (a) and rotated (b) joint trajectories along time.

with

$$\mathbf{T}_i = \begin{bmatrix} \cos\theta_i & 0 & \sin\theta_i \\ a & b & c \\ -\sin\theta_i & 0 & \cos\theta_i \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & -J^i_{0,x} \\ 0 & 1 & 0 & -J^i_{0,y} \\ 0 & 0 & 1 & -J^i_{0,z} \end{bmatrix}. \tag{3}$$

The output of this phase is the gait sequence G_O expressed with respect to the local and time-dependent system of coordinates where axis Z coincides with the walking direction. It is worth pointing out that, as a consequence of the transformation, the coordinates of each joint now represent a local variation with respect to the center of mass.

In Fig. 6 we compare the joints trajectories of a gait sequence before and after the proposed geometrical transformation. In Fig. 6-(a) the joints are shown in the native Kinect reference system: in the plot we selected the top-view (xz plane) as an example; the green markers refer to the central joints of body (head, shoulder center, spine and hip center), while red and blue ones refer to left and right body joints respectively (arms and lags). In the proposed example the subject is drifting to the right with respect to the sensor optical center. Fig. 6-(b) shows the values of $J^i_{k,X}$ computed by (2) as a function of the distance from the sensor ($\delta_i = d((0,0,0), J^i_0)$): it can be noted that $J^i_{k,X}$ capture local variations of the left and right joints with respect to the center of mass, that follows a horizontal straight trajectory.

4 Gait Features Definition and Classification

In this section we define a rich set of gait features exploiting the skeletal joint coordinates. In particular we are interested in estimating both classic anthropometric indexes, such height, arm length, etc., and dynamic features that are representative of the walking action.

4.1 Features Extraction

In the following we define and discuss two kind of features, namely *Features of Relative Distance* between joints (*FoRD*) and *Features of Sway* of joint (*FoS*).

*FoRD*s are sum of distances between pairs of joints. Let ξ_l^i be an ordered list containing a subset of the 20 joints of Ξ_i. We define the corresponding feature as the sum of the distances between pair of consecutive joints by considering the order provided by the list:

$$F(\xi_l^i) = \sum_{k=1}^{|\xi_l^i|-1} d(\xi_l^i(k), \xi_l^i(k+1)) \tag{4}$$

where k represents the k-th element of the list. As an example, the set $\xi_1^i = \{J_4^i, J_5^i, J_6^i, J_7^i\}$ includes the joints of the left arm (from the shoulder to the hand) and $F(\xi_1^i)$ represents the overall arm length at time i. Each *FoRD* is represented by the temporal average of the corresponding feature defined as $\langle F(\xi_1^i) \rangle$, where $\langle \cdot \rangle$ denotes the temporal average.

In our analysis we consider the 27 *FoRD*s listed in Table 1. The features with label $l = 1, \ldots, 21$ represent anthropometric parameters; in particular the first 6 features estimate the most standard anthropometric measures, i.e. the length of left/right arm, left/right leg, torso, and height. The remaining features ($l = 22, \ldots, 27$) are related to distance between pairs of symmetrical joints (not directly connected by skeleton bones), e.g. distance between left and right elbows or knees. We introduce these features since the movement of such joints depend on the walking style; as an example we can conjecture that the average distance between the left and right knees is affected by the walking dynamic.

The second set of features that we propose, called *FoS*s, aims at characterizing the walking style in terms of joints' sway in lateral, vertical, and frontal directions, i.e. the X, Y, Z axes defined in Section 3.3. This choice is driven by two considerations. First, in our previous work [33] we showed that other classical gait cycle features, e.g. stride length and walking speed, are difficult to estimate using Kinect due to its limited depth range (allowing only 3 to 4 strides acquisitions per gait sequence). More importantly, sway has been recognized as an important characteristic for both balance and gait analysis in the biomedical literature [38,39].

We define 6 *FoS* features for each joint computing the temporal average values $\langle J_{k,X}^i \rangle$, $\langle J_{k,Y}^i \rangle$, $\langle J_{k,Z}^i \rangle$ and corresponding deviations along the 3 axes. In

Table 1: *FoRD* Gait Features (L = left, R = right, C = center).

Feature	Relevant joints	Description
$\langle F(\xi_1^i)\rangle$	$\xi_1^i = \{J_{4-7}^i\}$	L arm
$\langle F(\xi_2^i)\rangle$	$\xi_2^i = \{J_{8-11}^i\}$	R arm
$\langle F(\xi_3^i)\rangle$	$\xi_3^i = \{J_{12-15}^i\}$	L leg
$\langle F(\xi_4^i)\rangle$	$\xi_4^i = \{J_{16-19}^i\}$	R leg
$\langle F(\xi_5^i)\rangle$	$\xi_5^i = \{J_{3-0}^i\}$	torso
$\langle F(\xi_6^i)\rangle$	$\xi_6^i = \{J_{3-0}^i, J_{12-15}^i\}$	height
$\langle F(\xi_7^i)\rangle$	$\xi_7^i = \{J_4^i, J_5^i\}$	L shoulder-elbow
$\langle F(\xi_8^i)\rangle$	$\xi_8^i = \{J_5^i, J_6^i\}$	L elbow-wrist
$\langle F(\xi_9^i)\rangle$	$\xi_9^i = \{J_4^i, J_5^i, J_6^i\}$	L shoulder-elbow-wrist
$\langle F(\xi_{10}^i)\rangle$	$\xi_{10}^i = \{J_8^i, J_9^i\}$	R shoulder-elbow
$\langle F(\xi_{11}^i)\rangle$	$\xi_{11}^i = \{J_9^i, J_{10}^i\}$	R elbow-wrist
$\langle F(\xi_{12}^i)\rangle$	$\xi_{12}^i = \{J_8^i, J_9^i, J_{10}^i\}$	R shoulder-elbow-wrist
$\langle F(\xi_{13}^i)\rangle$	$\xi_{13}^i = \{J_{12}^i, J_{13}^i\}$	L hip-knee
$\langle F(\xi_{14}^i)\rangle$	$\xi_{14}^i = \{J_{13}^i, J_{14}^i\}$	L knee-ankle
$\langle F(\xi_{15}^i)\rangle$	$\xi_{15}^i = \{J_{12}^i, J_{13}^i, J_{14}^i\}$	L hip-knee-ankle
$\langle F(\xi_{16}^i)\rangle$	$\xi_{16}^i = \{J_{16}^i, J_{17}^i\}$	R hip-knee
$\langle F(\xi_{17}^i)\rangle$	$\xi_{17}^i = \{J_{17}^i, J_{18}^i\}$	R knee-ankle
$\langle F(\xi_{18}^i)\rangle$	$\xi_{18}^i = \{J_{16}^i, J_{17}^i, J_{18}^i\}$	R hip-knee-ankle
$\langle F(\xi_{19}^i)\rangle$	$\xi_{19}^i = \{J_2^i, J_1^i, J_0^i\}$	C shoulder-spine-hip
$\langle F(\xi_{20}^i)\rangle$	$\xi_{20}^i = \{J_4^i, J_2^i, J_8^i\}$	L-C-R shoulder
$\langle F(\xi_{21}^i)\rangle$	$\xi_{21}^i = \{J_{12}^i, J_0^i, J_{16}^i\}$	L-C-R hip
$\langle F(\xi_{22}^i)\rangle$	$\xi_{22}^i = \{J_5^i, J_9^i\}$	L-R elbow
$\langle F(\xi_{23}^i)\rangle$	$\xi_{23}^i = \{J_6^i, J_{10}^i\}$	L-R wrist
$\langle F(\xi_{24}^i)\rangle$	$\xi_{24}^i = \{J_7^i, J_{11}^i\}$	L-R hand
$\langle F(\xi_{25}^i)\rangle$	$\xi_{25}^i = \{J_{13}^i, J_{17}^i\}$	L-R knee
$\langle F(\xi_{26}^i)\rangle$	$\xi_{26}^i = \{J_{14}^i, J_{18}^i\}$	L-R ankle
$\langle F(\xi_{27}^i)\rangle$	$\xi_{27}^i = \{J_{15}^i, J_{19}^i\}$	L-R foot

place of standard deviation we propose to use the *Median Absolute Deviation* (MAD), that is a robust measure of the statistical dispersion. For the k-th joint the lateral MAD is thus defined as:

$$\mathrm{MAD}(J_{k,X}) = Q_2\left(\left|J_{k,X}^i - Q_2\left(J_{k,X}\right)\right|\right) \tag{5}$$

where $Q_2(\cdot)$ denotes the median (or second quartile) of a series of values. The MAD in the vertical and frontal directions (axis Y and Z) is computed analogously.

These statistical features are computed for all joints with the exception of J_0 that coincides with the coordinate system origin. Therefore, we extract a total of 114 *FoS* features.

To summarize, considering both *FoRD* and *FoS* features, we have defined $f = 141$ gait parameters.

4.2 Dimensionality Reduction and Classification

Our ultimate goal is to perform people identification based on the set of gait features defined in Section 4.1. For classification purpose we employ a standard approach based on dimensionality reduction followed by well-known supervised method, namely the *Support Vector Machine* (SVM).

Let us consider g gait sequences of S different subjects and extract from each of them a feature vector, composed by f features: all the acquired gait data can be arranged in $g \times f$ matrix M. Dimensionality reduction is a quite common step to avoid overfitting the classification model when coping with a high number of features. To this end here we apply *Principal Component Analysis* (PCA) to limit the dimension of the features space. In particular, as common practice, the values in the feature matrix M are first normalized in the range $[0, 1]$. Then, eigenvector multivariate analysis is used to project the data into a subspace retaining a fixed percentage τ of the overall data energy. The output of this phase is a reduced features matrix M_τ of size $g \times f_\tau$, with $f_\tau < f$.

Finally, SVM is applied on M_τ. A fixed percentage of the acquisitions of each subject is used for training, i.e. a subset of the rows of M_τ, and the rest for testing. SVM, that has been originally designed for binary classification, can be extended to our S-classes case using two approaches: *One Against All* (OAA) and *One Against One* (OAO). These methods map the multi-class problem onto a set of binary ones. The OAA decomposition transforms the multi-class problem into a series of S binary subtasks used to discriminate a given class from all the others. The OAO defines $S(S-1)/2$ binary sub-tasks aiming at discriminating every possible pair of different classes.

5 Datasets

In this section we introduce the datasets used in our experimentation. The first one (referred to as *KinectUNITO'13*) and the third one (referred to as *KinectUNITO'16*) have been acquired in our laboratory while the second one (referred to as *KinectREID*) is a dataset available on request [31]. All datasets have been acquired with Kinect For Windows v1.

KinectUNITO'13 dataset (available in [40]) has been acquired in a controlled environment, where subjects have been forced to follow a given path, e.g. in security applications. In this case people aware of the video acquisition were asked to follow a straight path along a corridor with diffuse and almost constant lighting. The sensor has been placed at a convenient location for maximizing the field of view allowing us to capture sequences of 3 to 4 uninterrupted strides. The dataset includes gait samples of $S = 20$ subjects, 12 males and 8 females aged from 25 to 70. For each subject we have collected 20 gait sequences, 10 frontal views and 10 rear ones for a total of $g = 400$ gait sequences.

KinectREID dataset has been used in [31] to experiment a re-identification method that exploits both gait and dressing color features. In spite of the previous one, this scenario is representative of an uncontrolled video surveillance environment. The dataset includes gait samples of 71 subjects acquired in a lecture hall, under different lighting conditions and three view points. For each view point there are two gait sequences (frontal and rear), for a total of 6 gait samples per subject. Furthermore, some of the individuals carry accessories like bags. The walking area is wider than the sensor depth range and therefore depth acquisition and skeletal data are available only for a subset of the acquired frames. In order to test the proposed algorithm we need to track at least one gait cycle, that on average takes about 1 s. As a consequence, we select for our experiments the subset of $S = 45$ subjects, whose gait sequences G_O turn to be longer than 1 s after the gait data processing phase described in Section 3.3.

KinectUNITO'16 dataset has been acquired in a big lecture hall, letting the subjects follow a semicircular path. In addition, for reproducing a more realistic uncontrolled surveillance scenario we also collect gain sequences where the subjects carry some accessory: a shoulder bag, a backpack and a smartphone, for a total of 4 different scenarios (with any accessory or with one of these accessories). For each of these 4 scenarios we collect 4 walking sequences, 2 from the frontal view and 2 from the rear one, for a total of 16 samples per subject. Such scenarios have been chosen for a particular reason: in fact, the Kinect sensor is set up only for recognizing people standing in front of the camera, any other configuration represents a challenge for the skeleton tracking capability. The major inaccuracies in skeleton acquisition takes place when the arms are partially occluded, as in rear poses; when something covers the arm, like a shoulder strap; or when the arm is bent for keeping something. We have added scenarios concerning all these conditions to the dataset for analyzing in detail how much serious is the performance loss with such obstacles. This dataset contains the gait sequences of $S = 10$ subjects, 8 males and 2 females aged from 30 to 50, and a total of $g = 160$ gait sequences. All these subjects were already included in the *KinectUNITO'13*, but these new samples have acquired three years later.

6 Experimental Results

In this section we will analyze the performance of the proposed method in terms of person identification accuracy and compare our results with other approaches. Our experimental validation has been performed using the datasets presented in Section 5. The proposed gait features have been acquired by a prototype developed in C# language using the Kinect developer toolkit. PCA and SVM have been performed with the STPRTool [41], a free statistical pattern recognition toolbox. The SVM model has been trained using *Sequential Minimal Optimization* (SMO) [42] and regularization constant $C = 10$; radial

Table 2: Identification rate $CMC(1)$ for $\tau = 0.9$ as a function of the axes considered for *FoS* features.

Axis	f	f_τ	OAA $CMC(1)$	OAO $CMC(1)$
XYZ	141	25	0.95	0.95
XY	103	20	**0.97**	**0.96**
XZ	103	22	0.95	0.93
YZ	103	18	0.94	0.92
X	65	16	0.94	0.93
Y	65	10	0.89	0.89
Z	65	14	0.93	0.89

basis function (RBF) and linear kernel have been selected for the OAA and OAO decompositions method, respectively.

The following experimental analysis is based on the Monte Carlo cross-validation tests using 100 trails with fixed percentages of training and testing set. The performance has been evaluated in terms of *Cumulative Matching Characteristic* (CMC) curves. The $CMC(r)$ is a cumulative function of the rank r and represents the probability of finding the true identity among the top r identities according to the SVM output. Therefore, $CMC(1)$ represents the identification rate if the SVM output is used directly to attribute the identity, whereas $CMC(r)$ represents the probability to find the correct identity in the set of the r most likely ones. It is worth pointing out $CMC(r)$ performance metric can be highly relevant for practical applications, e.g. during the police investigations in which it is preferred to have a likely set of few suspicious individuals rather than an unique culprit.

6.1 Gait Features Analysis

The first set of experiments has been worked out using the *KinectUNITO'13* for analyzing the performance achievable in a controlled gait acquisition setting.

In Table 2 the identification rates $CMC(1)$ yielded by the proposed method using both OAA and OAO classifications are shown as a function of the considered gait features. In these first experiments we used $\tau = 0.9$ for dimensionality reduction and 70% of the samples for SVM training. The first row in the table refers to the case when all the 141 *FoRD* and *FoS* features, defined in Section 4.1, are used for identification: these features are mapped onto $f_\tau = 25$ most relevant features by PCA before SVM classification, that yields an identification rate equals to 0.95 for both OAA and OAO. The following rows of the table are obtained limiting the number of *FoS* used for classification: in particular, we test all possible combinations of axes used to compute sways. It is worth pointing out that when computing only vertical and horizontal sways (axes XY) and omitting variations along the walking direction (axis Z) the SVM identification rate slightly improves with $CMC(1) = 0.97$ and $CMC(1) = 0.96$ for OAA and OAO, respectively. On the contrary, all other combinations of axes determine a performance degradation. According to this

first analysis we can conclude that the most discriminative *FoS* features are those computed in the vertical and horizontal directions; as a consequence, in the following experiments we will omit *FoS* computation along Z.

In the next experiments we compare the identification accuracy offered by gait features in the vertical and horizontal axes ($f = 103$ followed by PCA dimensionality reduction) with respect to a set of 14 gait features experimentally identified as highly discriminant in [33] without PCA. This set of manually selected features includes a subset of the biometric and gait features defined in Section 4. In particular we use the following biometric parameters: length of arms $\langle F(\xi_1^i)\rangle$, $\langle F(\xi_2^i)\rangle$, legs $\langle F(\xi_3^i)\rangle$, $\langle F(\xi_4^i)\rangle$ and torso ($\langle F(\xi_5^i)\rangle$), subject height $\langle F(\xi_6^i)\rangle$. As representative of gait dynamic, we include the mutual distance between left and right elbows and knees $\langle F(\xi_{22}^i)\rangle$,$\langle F(\xi_{25}^i)\rangle$ and a limited set of *FoS* features, namely the average values of sway of head $\langle J_{3,X}^i\rangle$, $\langle J_{3,Y}^i\rangle$ and knees $\langle J_{13,X}^i\rangle$, $\langle J_{13,Y}^i\rangle$, $\langle J_{17,X}^i\rangle$, $\langle J_{17,Y}^i\rangle$. In both cases we perform identification with SVM using 70% training set. In Fig. 7 we compare the confusion matrices yielded by the manual set of features (top row) with respect to the proposed approach based on PCA (bottom) for both OAA and OAO. The figure shows that the method proposed in this paper significantly reduces misclassified samples (dots outside diagonal in figures).

In Fig. 8 the CMCs of the proposed method with different levels of PCA reduction ($\tau = 0.95, 0.9, 0.8$, corresponding to $f_\tau = 30, 20, 10$) are compared against the CMC obtained with the fixed set of 14 gait features (errors bars are used to represent the 95% confidence interval of each estimated value). The figures in the top row refer to case when 70% of the samples are used for training, whereas the bottom row shows results obtained with 50% training set. It can be observed that the proposed solution based on PCA significantly outperforms the manual feature selection for all values of r. It can be also noted that setting $\tau = 0.9$ yields a good balance between dimensionality reduction and identification performance. Finally, it is worth pointing out that the identification accuracy remains above 0.95 even when the training set is reduced to 50% of the samples.

6.2 Sensitivity to Acquisition Settings

Clearly the sensor tracking accuracy potentially impacts on the identification performance of the proposed algorithm. In particular, the sensor tracking capability, that is based on depth map measurements, is reduced when the observed subject is far away. Furthermore the Kinect tracking method is particularly effective in recognizing frontal poses that are more typical in gaming applications. To investigate this phenomenon in the following we test the sensitivity of the identification accuracy to the position of the sensor with respect to the observed scene. Since, according to the previous results, the performance of OAA and OAO classification is very similar, here we only show the results achieved with the first one. In the following we consider PCA with $\tau = 0.9$ and 70% training set.

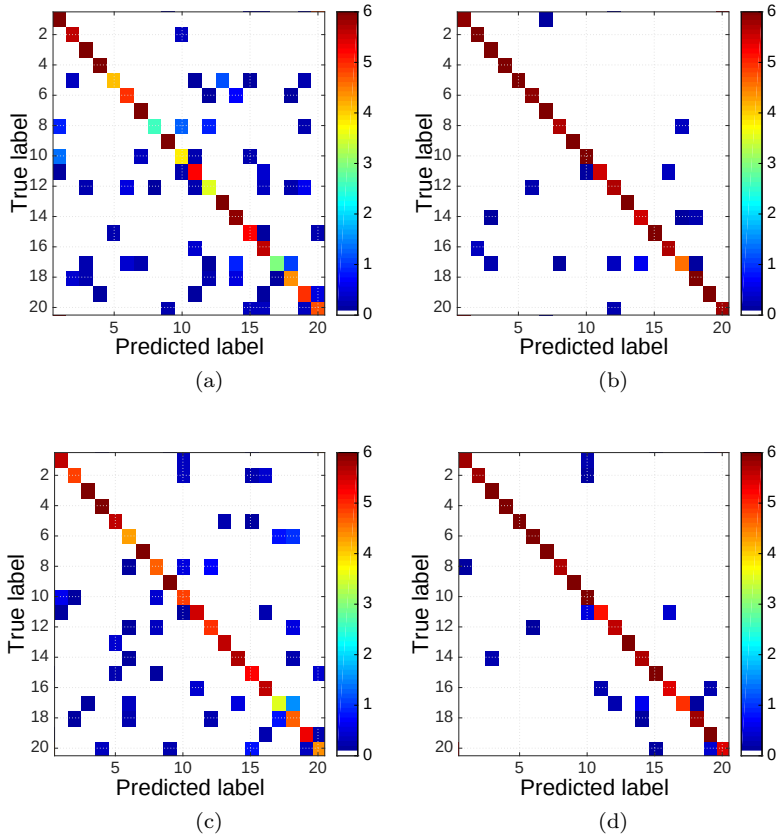


Fig. 7: Confusion matrices of OAA (a,c) and OAO (b,d) decomposition: classification with manual features selection (top) and proposed method with PCA (bottom), $\tau = 0.9$, training set 70%.

In the first test we use only frontal gait samples, i.e. 10 samples for subject using the *KinectUNITO'13* dataset. Out of these, 7 samples are used for training and the remaining 3 for testing. The same experiment is performed using only rear acquisitions. The obtained CMC is shown in Fig. 9-(a). It can be observed that the classification accuracy is only marginally affected by the type acquisition and that almost the same results are reported using front, rear and both views. These experiments show that the proposed solution is effective and robust for people identification and does not impose strict constraints on the acquisition direction, e.g. as systems based on face recognition that require a frontal view.

In the second test we split the gait sequence into two parts, corresponding to frames acquired when the subject is near (less than 2.6 m) and far (more than 3.2 m) from the sensor, respectively. The experimental results are shown

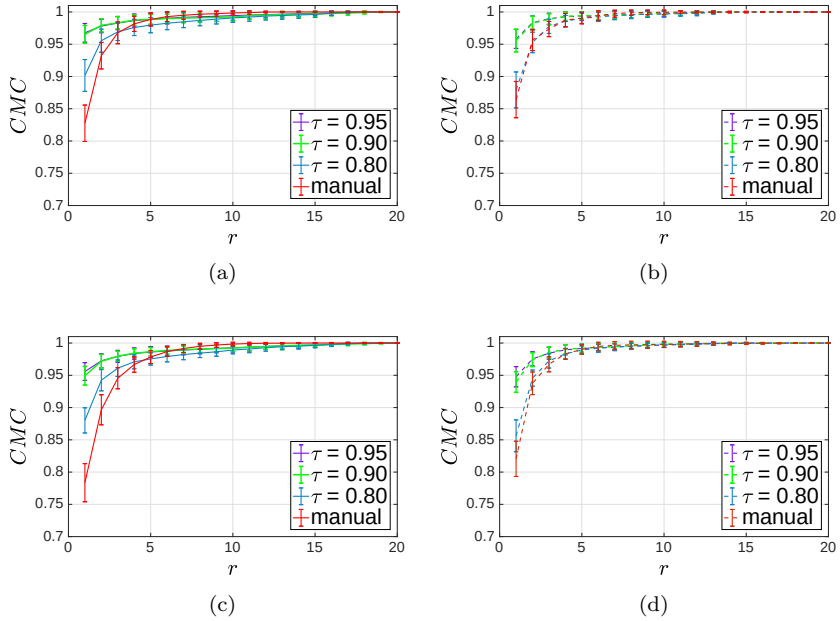


Fig. 8: CMC for OAA (a, c) and OAO (b, d) SVM classification: manual features selection vs proposed PCA with $\tau = 0.8, 0.9, 0.95$, training set 70% (top) and 50% (bottom).

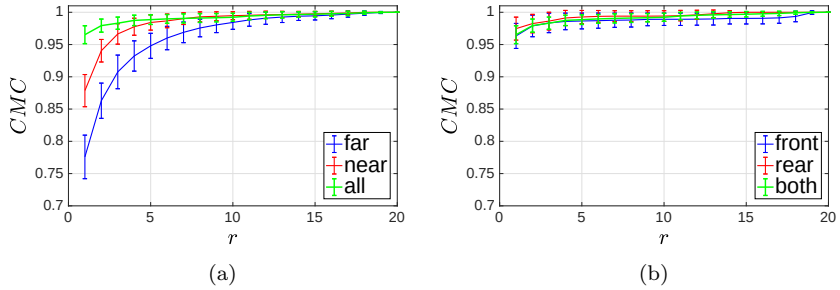
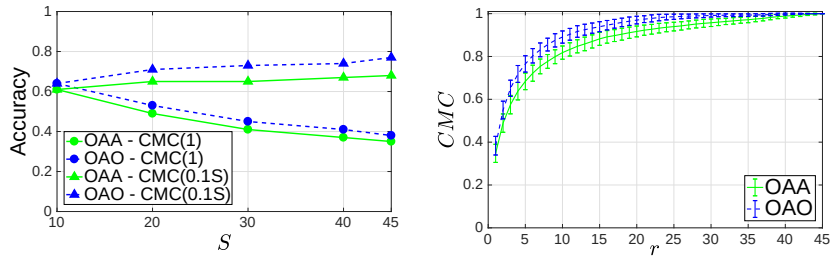


Fig. 9: Sensitivity to acquisition settings: CMC for OAA classification, $\tau = 0.9$, training set 70%, obtained using only front or rear (a) and using only near or far (b) acquisitions.

in Fig. 9-(b). We can notice that the accuracy decreases significantly with far acquisitions. This difference is likely to be caused by Kinect tracking error that is known to increase with distance from the sensor [43]. When using only near data the identification accuracy improves exhibiting CMC values above 0.85.



(a) (b)

Fig. 10: *KinectREID* dataset with $\tau = 0.9$ and 70% training: (a) CMC for $r = 1$ and $r = 0.1S$ as a function of S; (b) $CMC(r)$ in the case $S = 45$.

In any case the best performance (CMC above 0.95) is obtained when using all the available samples.

6.3 Identification in Uncontrolled Environment

Previous performance evaluation has shown that our method is effective for person identification in our controlled scenario. In the current section we study the performance obtained in an uncontrolled setting using *KinectREID* dataset. All the reported results are worked out using the proposed method with $\tau = 0.9$ (that yields in this case $f_\tau = 24$) and 70% SVM training. As discussed above *KinectREID* includes 45 subjects whose acquisitions are compatible with our algorithm.

In Fig. 10-(a) we show the $CMC(r)$ for $r = 1$ and $r = 0.1S$ as a function of $S = \{10, 20, 30, 40, 45\}$ for both OAA and OAO classification methods. The Monte Carlo cross validation in the cases with $S < 45$ has been worked out on subsets of subjects randomly selected out of the 45. By comparing the case $S = 20$ with the analogous experiment on *KinectUNITO'13* data in Fig. 8-(a,b) we can notice that, as expected, the second dataset represents a much harder scenario: in fact, $CMC(1)$ drops from about 0.95 to 0.6. Moreover, $CMC(1)$ decreases with the number of subjects to identify. Nonetheless, it is worth observing that the probability to have the correct identity in the top 10% ranking, i.e. $CMC(0.1S)$, increases with S reaching the best accuracy of 0.8 with $S = 45$. In Fig. 10-(b) the whole $CMC(r)$ is shown in the case $S = 45$ to better appreciate the accuracy of the obtained ranking in the toughest case we experimented. It can be noted that identification accuracy approaches 0.9 at $r = 10$. The performance degradation with respect to the one obtained on our dataset can be explained recalling the characteristics of *KinectREID*: first, the availability of only 6 gait sequences per subjects significantly limits the training set the SVM relies upon; moreover, data were acquired in uncontrolled setting with possible auto-occlusion (some body joints are not visible from

Table 3: Comparison with other methods: main characteristics and accuracy results.

Reference	**Data**	**Viewpoint**	**Method**	S	$CMC(1)$
[25]	Skeleton	Front	Similarity metric	20	0.92
[28]	Skeleton	Semicircular	SVM	20	0.96
[27]	Skeleton/Depth	Front/Back	Similarity metric	29	0.76
[31]	Skeleton/Depth/Color	Front/Back	Similarity metric	51	0.4
Proposed	Skeleton	Invariant	SVM	20	**0.97**
				45	0.4

some viewpoints) and suboptimal lighting conditions, e.g. strong back-light, where the infrared technology fails to provide precise tracking.

6.4 Comparisons With Related Works

The results obtained with our experimental analysis are very promising and our take away message is that gait features based on the sole skeletal tracking can be exploited for identity attribution. Unfortunately, direct comparisons with other results in the literature are difficult due to the absence of common reference datasets and testing conditions for gait analysis.

In Table 3 the main characteristics of some related works published in the last years are recalled along with the achieved identification results to allow a comparative analysis. Ahmed *et al.* in [25] report identification accuracy equal to 0.92 with $S = 20$ subjects. It is worth recalling that such results are obtained in a controlled setup with frontal walking only. In [28] Andersson *et al.* define 60 gait features, based on the angle between joints in the lower body part, and reach $CMC(1) = 0.96$ with $S = 20$ using SVM: it is quite similar to our results (0.97 for $S = 20$); nonetheless one must consider that in [28] more gait cycles have been acquired by tracking the subject along a semi-circular trajectory with the sensor moving on a spinning disk. Chattopadhyay *et al.* in [27] provide identification results on a larger set of subjects ($S = 29$); they assume a controlled setting where two depth sensors are used to jointly acquired the frontal view (from which skeleton data are estimated) and the back view (from which the subject silhouette is extracted). The achieved identification accuracy is 0.76. Finally, it is worth comparing the accuracy we achieved on *KinectREID* with $S = 45$, i.e. 0.4, with the results in [31]. Originally this dataset has been used in [31] for targeting a very different objective, i.e. re-identification over multi-camera tracking. To this end, they use biometric features (estimated form the skeleton joints) along with color features extracted from the RGB images. Their results show that color descriptors improve identification provided by the sole biometric parameters achieving $CMC(1) = 0.4$ with $S = 51$. Results show that our method is able to reach a similar accuracy on the same dataset without the use of color descriptors.

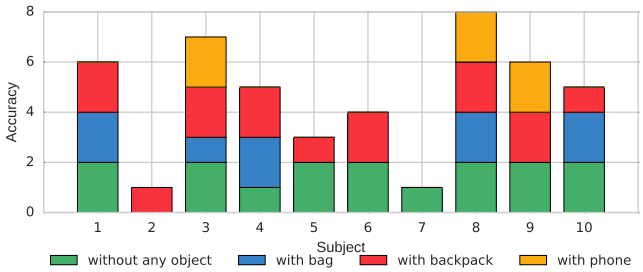


Fig. 11: Classification accuracy considering only front samples.

6.5 Identification Over Time

In this section we want to test the performance of the proposed method in terms of people identification over time. We use the samples in dataset *KinectUNITO'13* for training the SVM, while the samples in more recent dataset *KinectUNITO'16* are used in the testing phase. Our goal it to investigate if gait can be a distinctive biometric trait during the years: therefore we aim at recognizing an individual by comparing his current gait features with respect to the ones observed 3 years before. To this end, we consider only the *FoS* features yielding the best performance according to the results shown in Table 2, i.e. those related to the vertical and horizontal sways (axes XY): this amounts to a total of $f = 103$ gait features (including both *FoRD* and *FoS*).

As reported in Section 3, the Kinect experiences some difficulties in tracking the skeleton in rear pose. For this reason, in our first analysis we employ only the front samples of the two datasets. In this case, we train the SVM with 10 samples per subject in *KinectUNITO'13* and we consider four different testing conditions in *KinectUNITO'16*: people walking without any objects (a), with a bag (b), with a backpack (c), and while talking on the phone (d). For each subject we have two front samples per case, for a total of 8 gait sequences. We apply our method with $\tau = 0.9$, that reduces the feature dimensionality to 20, and SVM with linear kernel. The results of the classification task are shown in Fig. 11 where each bar represents the number of correct classifications per subject (over the 8 available testing cases), partitioned in different colors referring to the acquisition conditions (a)-(b)-(c)-(d). We can notice that 7 subjects have been correctly recognized in more than 50% of the cases, whereas subject 8 is always classified correctly. As expected, the classification task is more effective if the people's gait data are not perturbed by wearing accessories: in this case, both the available samples of 7 subjects have been correctly classified yielding an overall average accuracy of 0.8. It can be noted that wearing backpack (red color bar) does not influence classification much: 6 subjects are correctly classified with average accuracy that equals 0.75. On the contrary, by carrying a bag (blue bar) or a phone (yellow bar) one can

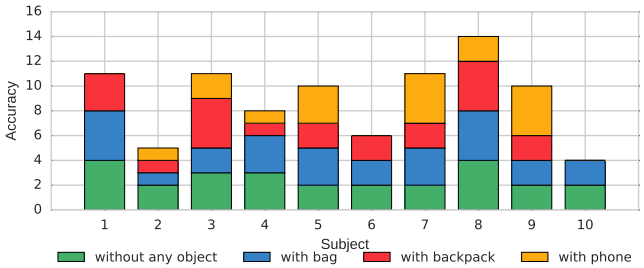


Fig. 12: Classification accuracy considering both front and rear samples.

easily deceive the classifier with an average accuracy that drops to 0.45 and 0.3 respectively. These results are clearly due to the different impact of the tested accessories on gait dynamic, with the backpack, that does not modifies arm and shoulder posture, interfering the least.

In the second classification experiment we also consider the rear poses in the two datasets: the training set now comprises 20 samples per subject, and the testing set contains 16 samples per subject, 4 for each acquisition condition (a)-(d). The experimental results obtained in this second classification task are shown in Fig. 12. If we analyze acquisition conditions separately, we can notice that as expected the best classification results are obtained in condition (a) without accessories with average accuracy of 0.65; nonetheless, this performance is significantly lower than the value of 0.8 obtained using only front view. With backpack, the accuracy decreases from 0.75 to 0.53: this is likely to be caused by the occlusion determined by the backpack on rear poses. On the other hand, it can be noted that the classification accuracy when an individual is carrying a bag or is making a phone call improves by exploiting rear views: in fact 5 subjects with a smartphone and 9 with a bag have been correctly classified at least 2 times over 4. When referring to the bag case the average accuracy increases to 0.65 from 0.45 reported in the previous experiment.

To sum up, our experiments show that gait features can be considered reliable biometric characteristic over the period of 3 years that we considered, with an accuracy as high as 0.8 reported on our dataset. Nonetheless, we have also shown that gait recognition in presence of perturbing accessories may easily become critical.

7 Conclusion

An automatic system for people recognition based on gait analysis has been proposed. Our approach turns out to be very effective for many reasons. The most informative frames of a gait sequence are detected automatically enforcing a set of constraints on the 3D skeletal data provided by the Kinect sensor. Gait analysis is based on a rich set of gait features that are made invariant with respect to the relative positioning between the sensor and the tracked

subject. Two classes of features are defined: the *Features of Relative Distance* between joints that extend classical biometric parameters, and the *Features of Sway* of joint that capture gait dynamic. The biometric recognition accuracy provided by the proposed gait features has been tested developing a people identification system based on PCA and supervised classification. The experiments have been performed using two acquisitions settings, namely a controlled and un-controlled setup (i.e. no preset path, rear poses and people carrying objects). Moreover, another objective of this study is to understand if gait can be considered an invariant biometric trait over years in our lifetime. Our experimental analysis shows that the proposed method is robust to acquisition settings and achieves very competitive identification accuracy with respect to the state of the art. In addiction, gait allows to recognize a person even after years, or at least after few years, making it compliant to forensic and security applications: just think of those situations where the perpetrator of crime has been observed through surveillance cameras years before trial. Moreover, we have also observed the impact of accessories, e.g. bag, phone, that significantly interfere with the gait dynamic making the classification task very tough. Further research in these cases is needed to overcome the limitations of the proposed system that is only partially able to exploit non frontal views. On the other side, the limited depth range of Kinect makes our method applicable only in specific real cases. A possible application could be identification in security check point in airports or banks, where the individual under inspection is alone in a limited space; as an example our method could be applied effectively to people passing through metal detector during airport security checks.

Other future works include the exploitation of the proposed gait features in security applications such as access control, people counting and identification, even in crowded scenario. Moreover, we envisage potential applications to other sectors such as health and aging where gait analysis can play a major role for identifications of pathologies and frailty conditions.

References

1. J. Ashbourn, *Biometrics - advanced identity verification: the complete guide.* (Springer, 2002)
2. S. Liao, A.K. Jain, S.Z. Li, IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(5), 1193 (2013)
3. A.K. Jain, A. Ross, S. Prabhakar, IEEE Transactions on circuits and systems for video technology **14**(1), 4 (2004)
4. R. Bolle, S. Pankanti, *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society* (Kluwer Academic Publishers, Norwell, MA, USA, 1998)
5. J.E. Boyd, J.J. Little, in *Advanced Studies in Biometrics* (Springer, 2005), pp. 19–42
6. L.F. Liu, W. Jia, Y.H. Zhu, in *Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence*, *Lecture Notes in Computer Science*, vol. 5755, ed. by D.S. Huang, K.H. Jo, H.H. Lee, H.J. Kang, V. Bevilacqua (Springer Berlin Heidelberg, 2009), pp. 652–659. DOI 10.1007/978-3-642-04020-7_70
7. R. Cucchiara, C. Grana, A. Prati, R. Vezzani, in *IEE Proceedings of Vision, Image and Signal Processing* (2005)

8. M. Goffredo, I. Bouchrika, J.N. Carter, M.S. Nixon, Multimedia Tools and Applications **50**(1), 75 (2010). DOI 10.1007/s11042-009-0378-5
9. P. Allard, *Three-dimensional analysis of human locomotion.* International Society Biomechanics series (Wiley, 1997)
10. P.K. Larsen, E.B. Simonsen, N. Lynnerup, in *Proc. Videometrics IX.*, vol. 6491 (2007), vol. 6491
11. I. Bouchrika, M. Goffredo, J. Carter, M. Nixon, Journal of Forensic Sciences **56**(4), 882 (2011)
12. J. Wang, M. She, S. Nahavandi, A. Kouzani, in *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on* (IEEE, 2010), pp. 320–327
13. S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, K.W. Bowyer, IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(2), 162 (2005). DOI 10.1109/TPAMI.2005.39
14. L. Wang, T. Tan, H. Ning, W. Hu, Pattern Analysis and Machine Intelligence, IEEE Transactions on **25**(12), 1505 (2003)
15. J. Han, B. Bhanu, Pattern Analysis and Machine Intelligence, IEEE Transactions on **28**(2), 316 (2006)
16. W. Kusakunniran, IEEE Transactions on Information Forensics and Security **9**(9), 1416 (2014)
17. D. Muramatsu, Y. Makihara, Y. Yagi, IEEE transactions on cybernetics **46**(7), 1602 (2016)
18. T. Connie, M.K.O. Goh, A.B.J. Teoh, IEEE transactions on cybernetics (2016)
19. R. Urtasun, P. Fua, in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on* (IEEE, 2004), pp. 17–22
20. I. Bouchrika, M.S. Nixon, in *Computer vision/computer graphics collaboration techniques* (Springer, 2007), pp. 150–160
21. F. Tafazzoli, R. Safabakhsh, Engineering applications of artificial intelligence **23**(8), 1237 (2010)
22. S.U. Jung, M.S. Nixon, IEEE Transactions on Information Forensics and Security **7**(6), 1802 (2012)
23. Y. Zhang, G. Pan, K. Jia, M. Lu, Y. Wang, Z. Wu, IEEE transactions on cybernetics **45**(9), 1864 (2015)
24. J. Preis, M. Kessel, M. Werner, C. Linnhoff-Popien, in *Proceedings of the First Workshop on Kinect in Pervasive Computing* (2012)
25. F. Ahmed, P.P. Paul, M.L. Gavrilova, The Visual Computer pp. 1–10 (2015)
26. P. Chattopadhyay, S. Sural, J. Mukherjee, IEEE Transactions on Information Forensics and Security **9**(11), 1843 (2014)
27. P. Chattopadhyay, S. Sural, and J. Mukherjee, Pattern Recognition Letters **63**, 9 (2015)
28. V. Andersson, R. Araujo, in *Proceedings of the Twenty-Ninth Association for the Advancement of Artificial Intelligence Conference, AAAI* (2015)
29. K. Yang, Y. Dou, S. Lv, F. Zhang, Q. Lv, Journal of Visual Communication and Image Representation (2016)
30. R. Satta, F. Pala, G. Fumera, F. Roli, in *8th International Conference on Computer Vision Theory and Applications (VISAPP 2013)* (Barcelona, Spain, 2013)
31. F. Pala, R. Satta, G. Fumera, F. Roli, IEEE Transactions on Circuits and Systems for Video Technology **8215**(MARCH), 1 (2015). DOI 10.1109/TCSVT.2015.2424056
32. E. Gianaria, N. Balossino, M. Grangetto, M. Lucenteforte, in *Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on* (2013), pp. 440–445. DOI 10.1109/MMSP.2013.6659329
33. E. Gianaria, M. Grangetto, M. Lucenteforte, N. Balossino, in *Biometric Authentication* (Springer, 2014), pp. 16–27
34. E. Gianaria, M. Grangetto, N. Balossino, in *International Conference on Image Analysis and Processing* (Springer, 2017), pp. 648–658
35. Z. Zhang, MultiMedia, IEEE **19**(2), 4 (2012)
36. K. Khoshelham, S.O. Elberink, Sensors **12**(2), 1437 (2012). DOI 10.3390/s120201437
37. M.a. Livingston, J. Sebastian, Z. Ai, J.W. Decker, 2012 IEEE Virtual Reality (VR) **298**(0704), 119 (2012). DOI 10.1109/VR.2012.6180911

38. J. Hegeman, E.Y. Shapkova, F. Honegger, J.H. Allum, Journal of Vestibular Research **17**(2), 75 (2007)
39. L.J. Janssen, L.L. Verhoeff, C.G. Horlings, J.H. Allum, Gait and Posture **29**(4), 575 (2009)
40. KinectUNITO gait dataset. URL `http://www.di.unito.it/~gianaria/project_gait.html`
41. V. Franc, V. Hlavác, Prague, Czech: Center for Machine Perception, Czech Technical University (2004)
42. B. Schölkopf, C.J. Burges, *Advances in kernel methods: support vector learning* (MIT press, 1999)
43. K. Khoshelham, S.O. Elberink, Sensors **12**(2), 1437 (2012)